Effects of *Banco de la Republica*'s Communication on the Yield Curve

Luis F. Melo-Velandia, Juan J. Ospina-Tejeiro, Julian A. Parra-Polania[1]

Banco de la República de Colombia

Abstract

We analyze the effect on the yield curve of Banco de la Republica's communication through two specific outlets, the minutes of the monetary policy meetings and the inflation reports during the period 2011-Q2 to 2018-Q4. We extract numeric information from the inflation reports' fan charts, and narrative information -using Latent Dirichlet Allocation, a computational linguistics tool- from the text of both outlets. We use an event-study approach to analyze the impact on four specific maturities: one-year spot, three-year forward, five-year forward and five-year ahead five-year forward rates. We find no evidence that numeric information has any effect on market yields. Regarding narrative variables we find that (i) for the inflation report, there is a significant effect on just two yields (one-year spot and five-year forward), and (ii) for the minute, there is a significant effect on all yields. We believe that these results may be explained by the publication lag of the inflation report during the period of analysis.

1. Introduction

Nowadays, it is widely considered that monetary policy can be more effective if it is credible and predictable. When central bank announcements are credible, it is easier to coordinate market expectations so that they contribute to achieving the targets set by the monetary authority. Consequently, communication becomes an essential instrument in the conduct of monetary policy (e.g. Woodford, 2005 and Blinder et al., 2008).

There exist various outlets for central bank communication, inter alia, press conferences, minutes, reports as well as public speeches of the members of the monetary policy committee. Numeric and narrative information is delivered through these outlets.

Given the relevance of communication, each central bank is likely interested in following up and assessing the corresponding strategy in order to determine whether or not the abovementioned different outlets disseminate information with clarity and in such a way that they contribute to the effectiveness of monetary policy, considering the particularities of the economy and the cultural context within which it operates.

[1] We thank Prof. José de Gregorio, Prof. Michael McMahon, Dr. Mario González and participants at the 2020 Workshop of the BIS-CCA Research Network for very helpful comments. We are especially grateful to Fredy Castañeda for his remarkable assistance and also to Daniela Machuca and Brandon Fino who provided assistance during specific parts of this project. All remaining errors are our own. The opinions expressed in this paper are those of the authors and not necessarily reflect the views of Banco de la República nor its Board of Directors.

Regarding some empirical evidence about the impact of central bank communication, previous literature has found evidence that, first, it has an effect not only on short-term market interest rates but also on the long-term ones and, second, the importance of the different channels that explain such an effect varies according to maturity (e.g. Gürkaynak et al., 2005; Cieslak and Schrimpf, 2019; Hansen et al., 2019a).

The present paper intends to contribute to the analysis of Banco de la Republica's (Colombian Central Bank) communication by studying the effect on the yield curve of the delivery of information to the public through two different outlets, the *minutes* of the monetary policy committee's meetings and *Inflation Reports*.

Some previous papers have analyzed the effect of monetary policy communication for the Colombian case. Castro (2012), using weekly data estimates a Fully Modified OLS and finds that only the speeches of members of the Board of Directors have a significant effect on some yield spreads ([7 years-1 year], [7 years-6 months], [5 years-1 year] and [5 years-6 months]), after controlling for other variables such as risk premium changes, macroeconomic news and policy rate changes. Arango et al. (2017) use text mining tools (in particular, Latent Semantic Analysis –LSA-) and a structural VAR and find that monetary policy press releases have an impact on (monthly) inflation expectations. Anzoategui Zapata and Galvis (2019) measure surprises in central bank communication through daily changes in the one-month IBR[2]. Then, using EGARCH models they find that (i) monetary policy communication has important effects on the (daily) returns of public debt securities and, (ii) in this respect, the minutes of the monetary policy meetings are the most significant outlet (compared to the inflation reports and press releases).

Our paper combines, as suggested by Hansen et al. (2019a), an event study approach and regressions that include numeric forecast information and signals from unstructured text data. The methodology allows not only to determine whether or not there is a significant impact of central bank's communication on the yield curve but also to find the specific types of information (within the minutes and the inflation reports, separately) that explain such an impact.

As explained by Bholat et al. (2015), analyzing numerical information has been a more common exercise in economics but, recently, other tools are being used to incorporate the analysis of narrative information. To extract narrative signals from both the minutes and the inflation reports, we use computational linguistics tools. Specifically, we use a model known as Latent Dirichlet Allocation (LDA) to try to identify the latent (hidden) topics present in the documents as a proxy of their narrative evolution. In this method, documents and words are assigned probabilistically to different topics (unlike LSA, where words -and documents- are assigned to topics with probability one), allowing for greater semantic flexibility in the analysis. By means of LDA we arrive at an 8-topic model to describe both the minutes and the inflation reports.

Our main results are obtained using (daily) data during the period 2011-Q2 to 2018-Q4 for four specific maturities: the one-year spot rate, the three-year forward rate, the five-year forward rate, and the five-year ahead five-year forward rate.

[2] IBR is the Spanish acronym for Reference Banking Indicator. It reflects the interest rate that banks expect to offer or to ask for funds in the interbank market.

We find no evidence that numeric information, namely the inflation and output growth forecasts that are contained in the inflation reports, has an impact on those market yields. However, regarding narrative signals, we find that: (i) for the case of the inflation reports, there is evidence that they have a significant effect on just two of the four rates analyzed (the one-year spot rate and the five-year forward rate), and (ii) for the case of the minutes, there is evidence that they have a significant impact on all the four market rates.

We think that an important element in explaining the abovementioned results is the fact that, during the analysis period, the inflation report was usually published one week after the publication of the corresponding minute and two to three weeks after the corresponding monetary policy meeting. With this lag it was unlikely that the reports were providing the market with new, unexpected information.

We also find that topics two (domestic demand and economic sectors) and four (inflation: target and inflation expectations) are the most relevant ones across yields, in the case of the minutes.

The remainder of this paper is organized as follows. Section 2 briefly describes the role and timing of the minutes and the inflation reports within the monetary policy communication process in Banco de la Republica. Section 3 describes data and the methodology. Section 4 presents results. Section 5 concludes.

2. Minutes and Inflation Reports as part of Banco de la Republica's communication

As aforementioned, we study the effect on the yield curve of the delivery of information to the public through the minutes of the monetary policy committee's meeting and the inflation reports.[3] The former one is elaborated by the Board of Directors. It explains the monetary policy decision and summarizes the main arguments in the deliberation process.

During most of our period of analysis (until 2017) **the minutes were published every month, about two weeks after the corresponding meeting**. Since 2018 the number of monetary policy meetings was reduced from 12 to 8 per year[4] and the lag between the meeting and the publication of the corresponding minute was reduced to one week. A more concise explanation of the decision has been previously provided through the press release and the press conference which take place right after the monetary policy meeting.

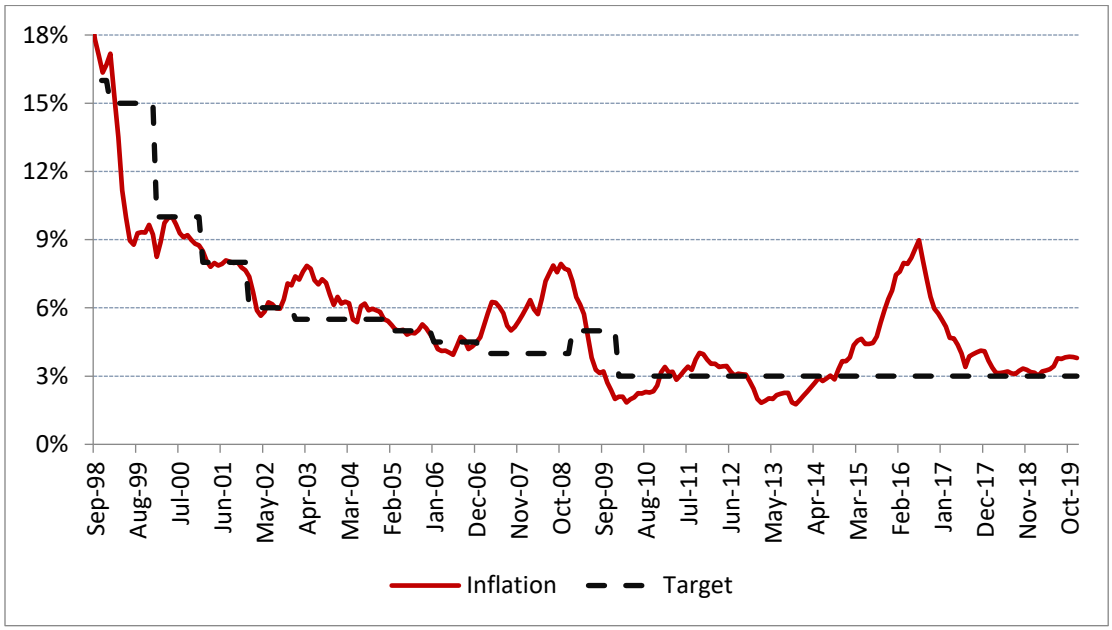
The inflation report, instead, is elaborated by the staff and contains the technical information and macroeconomic analysis that are presented to the Board of Directors to serve as an input for the monetary policy deliberation. It includes the staff's macroeconomic projections. **The inflation report is published quarterly**[5] and during the entire period of analysis (2011-2018) **it had a**

[3] The monetary policy committee in Colombia is usually referred to as the *Board of Directors* which consists of seven members with one vote each: the Minister of Finance, five full-time members and the General Manager (i.e. Governor) of the Bank, who is appointed by the other members.
[4] January, March, April, June, July, September, October and December.
[5] After the January, April, July and October monetary policy meetings.

publication lag of about two to three weeks with respect to the corresponding monetary policy meeting (and hence, it was published usually one week after the corresponding minute).[6]

The minutes and the inflation reports have been published since Jun-2007 and Jan-1999, respectively; however, our period of analysis is shorter due to data availability: inflation and output growth projections at one- and two–year horizons have been regularly released since the first inflation report of 2011.

Figure 1: Inflation and inflation target in Colombia



Our period of analysis corresponds to years during which, after having reduced inflation significantly, Banco de la Republica has decided to keep its annual inflation target equal to the long-run inflation target (3%). With the abandonment of the crawling peg regime, in 1999, Colombia became a full-fledged inflation targeting country. Inflation fell from levels around 18% in 1998 to around 4% in 2006. During 2007 and 2008, increments in the international price for oil and agricultural raw materials raised food and overall inflation (up to levels close to 8%). In 2009, partly because of the effects of the global financial crisis, inflation fell to 2% and, since 2010, Banco the la Republica has set the annual inflation target to 3%. Strong transitory supply shocks (climate impacts of El Niño and nominal depreciation) increased inflation up to levels of 7 and 9% in 2015 and 2016, but it then returned to 4 and 3% in 2017 and 2018 (Figure 1).

[6] Currently (2020), the minute is still published eight times per year and the inflation report is published quarterly; however, since October 2019, the publication lag of both was reduced to just one working day with respect to the monetary policy meeting.

3. Data and Methodology

To study the effect of the publication of Banco de la Republica's minutes and inflation reports on the yield curve, we follow some similar procedures as those applied by Hansen et al. (2019b) to analyze the case of the Bank of England's inflation report. We use an event-study approach for daily data during the period 01/05/2011 to 31/12/2018 (which includes the publication of 89 minutes and 32 inflation reports). The assumption behind the application of this approach is that changes of market interest rates on days of publication of either the minutes or the inflation reports are systematically caused by the news contained in such publications. Other causes would be specific (to just one or a few days) and non-systematic.

We estimate zero coupon yields using the Nelson-Siegel model with parameters obtained from Precia (a valuation prices provider in the Colombian Stock Exchange) for four different maturities: **the one-year spot rate (i_1), the three-year forward rate (f_3), the five-year forward rate (f_5), and the five-year ahead five-year forward rate ($_5f_5$)**.[7]

3.1. Numeric Information

This type of information is collected from the inflation reports only. The minutes of monetary policy meetings are not a systematic and primary source of inflation and output growth numeric projections: they do not provide inflation forecasts and only occasionally provide an output growth forecast (whose horizon may change in each occasion).

We extract 12 variables from numeric information (fan charts) in each inflation report. We denote the modal inflation and output growth forecast by π_t^i and g_t^i, respectively (where the superscript *i=12,24* indicates the forecast horizon, in months, and the subscript indicates the day of the inflation report publication). We also denote the (upper and lower) limits of the 90% probability interval around the mode by $Max[.]$ and $Min[.]$. Finally, we denote the surprise in each variable, relative to market expectations, by $Srp[.]$ (the next paragraph provides more details about how we construct these surprises). The 12 numeric-information variables are: four modal variables, $Srp[x_t^i]$, for *i=12, 24* and *x= π, g*; four dispersion measures, $Srp[Max[x_t^i] - Min[x_t^i]]$, for *i=12, 24* and *x= π, g*; and four asymmetry measures, $Srp[(x_t^i - Min[x_t^i]) - (Max[x_t^i] - x_t^i)]$, for *i=12, 24* and *x= π, g*.

Surprises are calculated in the following way. For π_t^{12} (the entire period of analysis) and π_t^{24} (during 2015-2018), as the absolute value of the difference between these modal forecasts and the median market inflation expectations for Colombia taken from Banco de la Republica's monthly expectations survey. For π_t^{24}, during 2011-2014 -since there is no information from the abovementioned survey-, as the absolute value of the difference between the modal forecast and the one in the previous inflation report. For g_t^{12}, as the absolute value of the difference between the modal forecast and the output growth market expectation for Colombia taken from the Latin Focus report (Consensus Forecast). For g_t^{24} and all the other variables (dispersion and asymmetry measures), as the absolute value of the difference between the forecast and the corresponding one in the previous inflation report.

[7] As in the case of Hansen et al. (2019b) we use nominal rather than real rates due to problems of availability and reliability of short-run real rates.

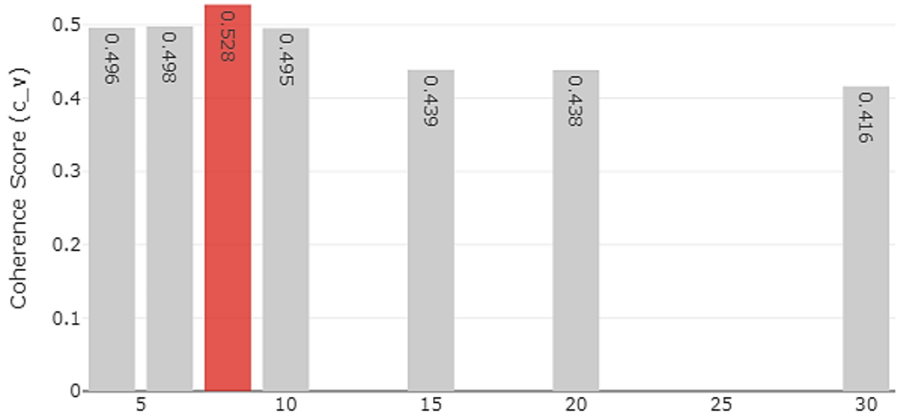
The whole set (vector) of variables extracted from numeric information is denoted by $\boldsymbol{Num}$.

3.2. Narrative Information

Narrative information is collected from both the minutes and the inflation reports. To this purpose, and as mentioned in the introduction, we use Latent Dirichlet Allocation (LDA) to try to identify the latent topics in the documents. It is a Bayesian tool in which each document is modeled as a combination of an underlying set of topics.[8]

After a standard textual pre-processing[9], we train the model using all documents (17.833 paragraphs in total, 13.920 (78%) from the inflation reports and 3.963 (22%) from the minutes)[10], and therefore we assume that the minutes and the reports refer to the very same topics.[11] Then, using the selected model, we analyze the minutes and the reports separately, so that we can estimate a topic distribution for each outlet. All the analysis is conducted using the documents in their original language (Spanish).

Figure 2: Coherence Score for different number of topics, *k= 4, 6, 8, 10, 15, 20* and *30*.



To select the appropriate number of topics (*k*), and hence the appropriate model, we calculate a coherence score (C_v) for different cases: *k= 4, 6, 8, 10, 15, 20* and *30.* Intuitively, the coherence score intends to measure the quality of the resulting topics based on the hypothesis that words

[8] More details on this tool can be found in Blei et al. (2003)

[9] Which consists of steps such as removing punctuation, non-alphabetic terms, the so-called *stop-words* (e.g. articles, preposition)… more details about analytical pre-processing can be found in Bholat et al. (2015).

[10] For training purposes we use the minutes and the inflation reports published during the period Jun-2007 to Dec-2018. The initial date of such period coincides with the moment in which the Colombian central bank begins to publish the minutes.

[11] To appropriately estimate an LDA model it is ideal to have many thousands of documents and therefore we use the minutes and the inflation reports together to estimate a unique model. However, to check if our assumption is reasonable, we estimate two separate models and find that their topics are very similar (results are available upon request).

with similar meaning tend to co-occur within a similar context.[12] From our coherence score calculations, we obtain that the optimal number of topics is 8 (see Fig. 2).

The LDA tool does not provide you with labels for topics, and hence their interpretation is entirely left to the researcher. In our case, based on both the most relevant words associated to each topic and exclusiveness of words for each topic, we label the eight resulting topics as follows:[13]

Topic 1: Interest rates and monetary policy

Topic 2: Domestic demand and economic sectors

Topic 3: Balance of trade and commodities

Topic 4: Inflation: target and expectations

Topic 5: Inflation: food and regulated items

Topic 6: Credit

Topic 7: External conditions

Topic 8: Macroeconomic projections

We extract 32 variables from narrative information, 16 from the minutes and 16 from the inflation reports. These variables correspond to the participation (or weights) of each of the eight topics within the topic distribution of the documents (minutes/reports), which we denote by w_t^j for *j=1, 2, … 8*, and the absolute value of the corresponding change, $\left|w_t^j - w_{t-1}^j\right|$ for *j=1, 2, … 8*. The whole set (vector) of variables extracted from narrative information is denoted by $\boldsymbol{Nar^{min}}$ for the case of the minutes and by $\boldsymbol{Nar^{rep}}$ for the case of the inflation reports.

The figures in Appendix B show the weight of each topic within the inflation reports and the minutes, for the period Mar-2011 to Sep-2018. In both types of documents, domestic demand and economic sectors (topic 2) and macroeconomic projections (topic 8) have represented important proportions of the narrative information. However, in the case of the minutes, while the latter has recently gained relevance the former topic has lost it.[14]

3.3. Event-study Regressions

We start by constructing narrative shocks ($\boldsymbol{v_t}$) by filtering out variation in narrative information (16 variables) that is already captured by numeric information (or to international volatility only, in the case of the minutes):

$$Nar_{t,j}^{rep} = \eta_{0,j} + \boldsymbol{\eta_{1,j}} \boldsymbol{Num_t} + \eta_{2,j} VIX_t + v_{t,j}^{rep} \quad (1)$$

for *j*= 1, 2, … 16

[12] The algorithm segments the data into word pairs, calculates word pair probabilities and association measures between sets of words and then aggregates those measures.

[13] Table A1 in Appendix A shows the topics, average weights and their corresponding more relevant words. Exclusiveness of words is not presented here but it is available upon request.

[14] See Guio-Martinez et al. (2020) for a more detailed description of the Central Bank of Colombia's minutes and inflation reports using LDA (available in Spanish only).

$$Nar_{t,j}^{min} = \mu_{0,j} + \mu_{1,j} VIX_t + v_{t,j}^{min} \quad (2)$$

for *j*= 1, 2, … 16.

The *VIX* index is included to consider the effect of international market volatility.

We want to estimate the explanatory power of both numeric information and narrative shocks for each of the four market interest rates; however we have too many regressors, and therefore to prevent overfitting, as suggested by Hansen et al. (2019b), we use elastic net regression,[15] a regularization method, to select the three most relevant narrative shocks ($\boldsymbol{v_t^*}$) that have an impact on the part of yields that is not explained by numeric information.[16]

Finally, to estimate the impact of the selected narrative shocks ($\boldsymbol{v_t^*}$) on market yields we conduct the following regressions:

$$\left|\Delta y_t^{rep}\right| = \gamma_{0,y} + \boldsymbol{\gamma_{1,y}} \boldsymbol{Num_t} + \gamma_{2,y} VIX_t + \boldsymbol{\gamma_{3,y}} \boldsymbol{v_t^{*rep}} + \epsilon_{t,y}^{rep} \quad (3)$$

for y = i_1, f_3, f_5 and ${}_5f_5$

$$\left|\Delta y_t^{min}\right| = \delta_{0,y} + \delta_{1,y} VIX_t + \boldsymbol{\delta_{2,y}} \boldsymbol{v_t^{*min}} + \epsilon_{t,y}^{min} \quad (4)$$

for y = i_1, f_3, f_5 and ${}_5f_5$

where $\left|\Delta y_t^{rep}\right|$ is the absolute value of the daily change in the market interest rate y, observed on the day of the report publication (*t*) and $\left|\Delta y_t^{min}\right|$ is the same variable but observed on the day of the minute publication (*t*).[17]

4. Results

Table 1 reports the top three shocks that are selected for each maturity in the case of the inflation reports and Table 2 does the same for the minutes. For instance, the third column of Table 1 tells us that (the absolute value of) the changes of topic two (domestic demand and economic sectors), and four (inflation: target and expectations), *D2* and *D4*, and the level of the latter, *L4*, are the three most relevant narrative shocks to explain the part of the three-year forward rate that is not

[15] Elastic net is a regression method described by Zou and Hastie (2005) that penalises variables according to the magnitude of their estimated coefficients. See Appendix C.

[16] Appendix C provides more details about how we carry out the selection of the three most relevant narrative shocks. Unlike Hansen et al. (2019b) we select three rather than four shocks since we have 16 narrative variables while they have 60.

[17] Hansen et al. (2019b) also conduct the same analysis splitting yields into the expectations and the term premium components and find robust evidence in favor of the importance of the uncertainty channel (i.e. central bank communication affects long-run interest rates by providing news on uncertainty around economic conditions). We decompose yields into the same two components (using an affine model as proposed by Adrian et al. 2013) but find no evidence of the uncertainty channel or any other interesting outcome, and hence we do not present results split by components. In the case of narrative information, a likely explanation for this difference is the fact that our model produces eight general topics such that none can be exclusively associated to either the expected value of economic conditions or the variance/uncertainty of such conditions, unlike the ones in the 30-topic model used by Hansen et al. (2019b).

explained by numeric information. The fourth column of the same table reports that, as another example, the level of topic 4 was selected as a relevant narrative shock in 85,8% of the 5000-simulation bootstrap conducted for the case of the three-year forward rate. See Appendix C for more details on this procedure.

Table 1: Inflation reports, selected narrative shocks for each maturity

$\lvert\Delta i_1\rvert$		$\lvert\Delta f_3\rvert$		$\lvert\Delta f_5\rvert$		$\lvert\Delta_5 f_5\rvert$	
Topic	%	Topic	%	Topic	%	Topic	%
L5	99,7	D2	98,7	D2	99,3	L3	99,8
D8	99,1	L4	85,8	L3	98,1	D6	99,4
L1	98,6	D4	82,3	D6	98,0	L5	98,3

Note: L refers to the shock related to the topic's weight (within the entire topic distribution) and D to the shock related to (the absolute value of) the 1st difference of such a weight. The reported narrative shocks are those three that were selected the highest fraction of times across the 5000-simulation bootstrap (see Appendix C). Column "%" reports such a fraction of times, in percentage.

Table 2: Minutes, selected narrative shocks for each maturity

$\lvert\Delta i_1\rvert$		$\lvert\Delta f_3\rvert$		$\lvert\Delta f_5\rvert$		$\lvert\Delta_5 f_5\rvert$	
Topic	%	Topic	%	Topic	%	Topic	%
L3	42,8	D7	74,4	L2	64,2	L2	76,2
D8	38,8	L2	30,0	D7	56,6	L4	37,6
L4	36,8	L4	27,8	L4	53,0	L5	37,0

Note: L refers to the shock related to the topic's weight (within the entire topic distribution) and D to the shock related to (the absolute value of) the 1st difference of such a weight. The reported narrative shocks are those three that were selected the highest fraction of times across the 500-simulation bootstrap (see Appendix C). Column "%" reports such a fraction of times, in percentage.

Regarding the inflation reports, the level of topic three (balance of trade and commodities) and the change of topic six (credit) are selected for both the five-year forward rate and the five-year ahead five-year forward rate. Furthermore, the change of topic two (domestic demand and economic sectors) is selected for the three-year forward rate and also for the five-year forward rate. It is also interesting to note that the level of topic five (inflation: food and regulated items) is selected for a short-run rate (the one-year spot rate) but also for a long-run rate (the five-year ahead five-year forward rate). However, to interpret these results it is also important to take account of the fact that, as shown below, the selected narrative shocks are found to be significant (at the 5% level of significance) only for the one-year spot rate and the five-year forward rate, in the case of the inflation reports.

In the case of the minutes, the shock related to the level of topic four (inflation: target and expectations) is selected for all maturities, and the level of topic two (domestic demand and economic sectors) is selected for all but the shortest one (Table 2).[18]

To check in a more complete and formal way whether different narrative shocks drive different rate maturities we use the whole set of shocks, arranged for each maturity according to the number of times they were selected in the bootstrap procedure (see Appendix C), and calculate Spearman correlation coefficients. Table 3 summarizes results for the inflation reports and Table 4 does the same for the minutes.

Table 3: Inflation reports, Spearman correlations of narrative shocks across yields

	$\lvert\Delta i_1\rvert$	$\lvert\Delta f_3\rvert$	$\lvert\Delta f_5\rvert$	$\lvert\Delta_5 f_5\rvert$
$\lvert\Delta i_1\rvert$	1,00			
$\lvert\Delta f_3\rvert$	0,16	1,00		
$\lvert\Delta f_5\rvert$	0,49*	0,41	1,00	
$\lvert\Delta_5 f_5\rvert$	0,32	-0,11	0,44*	1,00

Note: correlations are based on the rankings of the 16 narrative shocks according to the number of times they were selected in the bootstrap procedure (see Appendix C). Significance at the 10, 5 and 1% are denoted by *, ** and ***, respectively.

Table 4: Minutes, Spearman correlations of narrative shocks across yields

	$\lvert\Delta i_1\rvert$	$\lvert\Delta f_3\rvert$	$\lvert\Delta f_5\rvert$	$\lvert\Delta_5 f_5\rvert$
$\lvert\Delta i_1\rvert$	1,00			
$\lvert\Delta f_3\rvert$	0,23	1,00		
$\lvert\Delta f_5\rvert$	0,15	0,70***	1,00	
$\lvert\Delta_5 f_5\rvert$	0,46*	0,39	0,53**	1,00

Note: correlations are based on the rankings of the 16 narrative shocks according to the number of times they were selected in the bootstrap procedure (see Appendix C). Significance at the 10, 5 and 1% are denoted by *, ** and ***, respectively.

In the case of inflation reports, there is not robust evidence that there is correlation between the key narrative shocks across yields (Table 3). In the case of the minutes, there is stronger evidence of a positive correlation between the key shocks for the three-year forward rate and those for the five-year forward rate and between the latter ones and those for the five-year ahead five-year

[18] It may be also interesting to mention that although the level of topic three (balance of trade and commodities) is selected for the one-year spot rate only, when splitting yields into components, L3 is selected for all maturities in the term premium component and for all but the longest one in the expectations component (in the case of the minutes).

forward rate (Table 4). From the results, the evidence is not strong enough to state that there is a clear separation between the narrative shocks that drive long-run versus those that drive short-run market rates (in contrast to what is found by Hansen, 2019a, for the UK): at a significance level of 10%, there is evidence of a positive correlation between the narrative shocks that drive the one-year spot rate and those that drive the five-year forward rate, in the case of the inflation reports –Table 3-, and between the narrative shocks that drive the one-year spot rate and those that drive the five-year ahead five-year forward rate, in the case of the minutes –Table 4-.

Regarding the impact of numeric and narrative information on market yields, we follow a very similar procedure to the one in Hansen et al. (2019b). We separate the 12 numeric variables between four expectations variables (exp) –the modal forecasts- and eight uncertainty variables (unc) – the dispersion and the asymmetry forecasts-. We firstly run regressions with the VIX index and the expectations forecasts, then we add the uncertainty forecasts and finally we add the selected narrative shocks to see if each group has different impact on the market interest rates and if such impacts change according to maturities.

Table 5: Inflation Reports, effect of numeric information and selected narrative shocks on market interest rates

		Exp	*p-value*	*Add Unc*	*p-value*	*Add SNS*	*p-value*
$\lvert\Delta i_1\rvert$	R^2	0,183	0,398	0,365	0,732	**0,805**	0,016
	Additional R^2			0,182	0,785	**0,440**	0,001
	Partial R^2			0,223		0,693	
$\lvert\Delta f_3\rvert$	R^2	0,185	0,390	0,363	0,737	0,633	0,273
	Additional R^2			0,178	0,796	0,270	0,060
	Partial R^2			0,218		0,423	
$\lvert\Delta f_5\rvert$	R^2	0,158	0,498	0,329	0,819	**0,774**	0,034
	Additional R^2			0,171	0,832	**0,445**	0,002
	Partial R^2			0,203		0,664	
$\lvert\Delta_5 f_5\rvert$	R^2	0,114	0,686	0,286	0,898	0,559	0,486
	Additional R^2			0,171	0,853	0,273	0,090
	Partial R^2			0,193		0,382	

Note: Exp refers to expectations forecasts, Unc to uncertainty forecasts and SNS to selected narrative shocks. Additional and partial R^2 are related to the addition of regressors: uncertainty forecasts, at first, and then selected narrative shocks. The p-value refers to the F-test for the joint significance of the corresponding regressors (the whole set of regressors or only the additional ones in the corresponding regression). Results that are significant at the 5 % level of significance are in bold.

Table 5 summarizes the results about the joint explanatory power of the numeric and narrative shocks for the inflation reports and Table 6 does the same for the minutes. Remember that for the case of the minutes the only numeric variable considered is the VIX index.

With regard to numeric information from the inflation reports, there is no evidence (not even at high levels of significance) that it has impact on any of the four market interest rates. Neither expectations forecasts nor uncertainty forecasts have a statistically significant effect on those market yields (Table 5, columns 3-6). The same can be said in the case of the minutes for the VIX index (Table 6, columns 3-4).[19]

Table 6: Minutes, effect of the VIX index and selected narrative shocks on market interest rates

		VIX	*p-value*	*Add SNS*	*p-value*
$\lvert\Delta i_1\rvert$	R^2	0,000	0,853	**0,111**	0,041
	Additional R^2			**0,110**	0,020
	Partial R^2			0,110	
$\lvert\Delta f_3\rvert$	R^2	0,008	0,397	**0,111**	0,040
	Additional R^2			**0,103**	0,026
	Partial R^2			0,104	
$\lvert\Delta f_5\rvert$	R^2	0,000	0,922	**0,136**	0,014
	Additional R^2			**0,136**	0,006
	Partial R^2			0,136	
$\lvert\Delta_5 f_5\rvert$	R^2	0,000	0,895	**0,112**	0,039
	Additional R^2			**0,112**	0,018
	Partial R^2			0,112	

Note: SNS refers to selected narrative shocks. Additional and partial R^2 are related to the addition of SNS as regressors. The p-value refers to the F-test for the joint significance of the corresponding regressors (the whole set or only the addition of SNS in the corresponding regression). Results that are significant at the 5 % level of significance are in bold.

It seems unusual that revealing central bank forecasts has no impact on market interest rates; however, we believe that it may occur due to the publication lag of the inflation report during the period of analysis, that is to say, two to three weeks after the corresponding monetary policy meeting. As explained in Section 2, the inflation report contains the technical analysis that is presented by the staff to the Board of Directors to serve as an input for its deliberation. It is likely that one week after the minutes of the monetary policy meeting had been released and two to three weeks after the monetary policy decision had been announced and explained by means of the press release and the press conference, the inflation report were revealing no surprising information to the market.

Our results related to narrative information might be explained by the publication lag of these two outlets as well. At the 5% level of significance there is evidence that the selected narrative shocks

[19] We obtain similar results, and hence the same conclusion follows, when we conduct the analysis splitting yields into the expectations and the term premium components (see footnote 17). Those results are available upon request.

extracted from the inflation reports have impact only in the case of two market rates, the one-year spot rate and the five-year forward rate (Table 5, columns 7-8). Instead, at the same level of significance, in the case of the minutes there is evidence of such an impact on all the four market rates (Table 2, columns 5-6). It seems clear that surprising information can be extracted from the minute's text by the market, while there is not strong evidence of the same fact regarding the inflation report.[20] The publication lag of the inflation report with respect to the monetary policy meeting and the fact that the corresponding minute comes first, one week earlier, may be affecting the possibility of the report to provide the market with unexpected information.

5. Conclusion

This paper analyzes the effect of Banco de la Republica's communication on the yield curve. Particularly, it studies the impact of the delivery of information to the market through two different outlets, the minutes of the monetary policy committee's meetings and the inflation reports.

We extract two types of information, namely numeric and narrative. The former is obtained from output growth and inflation forecasts delivered through the inflation report's fan charts. Narrative information is extracted from the text of the minutes and the inflation reports by means of Latent Dirichlet Allocation (LDA), a computational linguistics tool which tries to identify the latent topics within a set of documents. We arrive at an 8-topic model to describe both outlets.

We then use an event-study approach (as suggested by Hansen et al. 2019b) to analyze the effect of the publication of those types of information on market interest rates. We study four specific maturities: the one-year spot rate, the three-year forward rate, the five-year forward rate, and the five-year ahead five-year forward rate. For the analysis we use data from 2011-Q2 to 2018-Q4.

With regard to numeric information extracted from the inflation reports, we find no evidence that it has an impact on market yields. Regarding narrative variables obtained by means of the aforementioned 8-topic model, we find that: (i) for the case of the inflation report, there is evidence that they have a significant effect on just two of the four rates analyzed (the one-year spot rate and the five-year forward rate), and (ii) for the case of the minute, there is evidence that they have a significant impact on all the four market rates.

We believe that these results may be explained by the publication lag of the inflation report during the period of analysis. Such a report was usually published one week after the publication of the corresponding minute and two to three weeks after the corresponding monetary policy meeting. With such a lag it was unlikely that the reports were providing the market with unexpected information. Since October 2019, both the minutes and the inflation reports are being published

[20] Since the VIX index and numeric information are not significant, we conduct the same regressions including the selected shocks only. The results are similar (at the 5% level of significance): in the case of the reports only the five-year forward rate is significant, while in the case of the minutes all maturities are. Also, since availability of numeric information was the reason to limit our period of analysis it seems reasonable to restrict our attention to narrative information only and work with longer time series. This was not a straightforward option as from 2007 to 2010 the market rate series display a systematically different behavior from the one during the original period of analysis (2011-2018); in particular, higher levels and significantly higher volatility (maybe related to Global Financial Crisis effects).

just one working day after the monetary policy meeting, and hence it is currently expected that more news is contained in both outlets. Besides being published in a more timely fashion, the new Inflation Report (now called Monetary Policy Report) is written in a more forward-looking way. It explicitly reports values for different assumptions, key parameters such as the neutral interest rate, and point forecasts; and also contains a qualitative comparison between the monetary policy rate expected by market analysts and the one expected by the bank's staff. The results of this paper support such changes. It remains to be seen if these changes would in turn modify the evidence presented here, should one repeat this study in the future.

We also find that topics two (domestic demand and economic sectors) and four (inflation: target and inflation expectations) are the most relevant ones across yields, in the case of the minutes. However, when considering the whole set of topics, there is no evidence that there is strong correlation between the key narrative variables that drive long-run versus those that drive short-run market rates.

REFERENCES

Adrian, Tobias & Crump, Richard K. & Moench, Emanuel, 2013. "**Pricing the term structure with linear regressions**," Journal of Financial Economics, Elsevier, vol. 110(1), pages 110-138.

Arango Luis E. & Javier Pantoja & Carlos Velásquez, 2017. "**Effects of the central bank's communications in Colombia. A content analysis**," Borradores de Economia 1024, Banco de la Republica de Colombia.

Anzoátegui Zapata, Juan Camilo & Juan Camilo Galvis, 2019. "**Efectos de la comunicación del banco central sobre los títulos públicos: evidencia empírica para Colombia**," Revista Cuadernos de Economía, Universidad Nacional de Colombia -FCE - CID, vol. 38(77), pages 337-364, July

Blei, David & Andrew Ng & Michael Jordan, 2003. "**Latent Dirichlet Allocation,"** Journal of Machine Learning Research, vol. 3, pages 993-1022.

Blinder, Alan S. & Michael Ehrmann & Marcel Fratzscher & Jakob De Haan & David-Jan Jansen, 2008. "**Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence**," Journal of Economic Literature, American Economic Association, vol. 46(4), pages 910-945, December.

Bholat, David & Stephen Hans & Pedro Santos & Cheryl Schonhardt-Bailey, 2015. "**Text mining for central banks**," Handbooks, Centre for Central Banking Studies, Bank of England, number 33.

Castro, Freddy , 2012. "**Señales de política monetaria y movimientos en la estructura a plazo de la tasa de interés en Colombia**," Archivos de Economía 009908, Departamento Nacional de Planeación.

Cieslak, Anna & Schrimpf, Andreas, 2019. "**Non-monetary news in central bank communication**," Journal of International Economics, Elsevier, vol. 118(C), pages 293-315.

Hansen, Stephen & McMahon, Michael & Tong, Matthew, 2019a. "**The Long-Run Information Effect of Central Bank Communication**," Journal of Monetary Economics, Elsevier, vol. 108, pages 185-202.

Hansen, Stephen & McMahon, Michael & Tong, Matthew, 2019b. "**The Long-Run Information Effect of Central Bank Communication**," Bank of England working papers 777, Bank of England.

Guío-Martínez, Daniela V. & Muñoz-Bravo, Germán A. & Ospina-Tejeiro, Juan J. & Parra-Polanía, Julián A., 2020. " **Description of the Minutes and Monetary Policy Reports using Computational Linguistics tools**," Borradores de Economia 1108, Banco de la Republica de Colombia (In Spanish).

Woodford, Michael, 2005. "**Central Bank Communication and Policy Effectiveness**," NBER Working Papers 11898, National Bureau of Economic Research, Inc.

Zou, Hui & Hastie, Trevor , 2005. **Regularization and variable selection via the elastic net.** Journal of the Royal Statistical Society, Series B 67 (2), pp. 301–320.

Appendix A

Topic	2	8	5	3	7	1	6	4
Label	**Domestic demand and economic sectors**	**Macroeconomic projections**	**Inflation, food and regulated items**	**Balance of trade and commodities**	**External conditions**	**Interest rates and monetary policy**	**Credit**	**Inflation target and expectations**
Average Weight	22.5%	15.8%	14.9%	12.0%	10.2%	9.1%	7.9%	7.6%
Words	growth	growth	prices	exports	united_states	interest_rate	credit	inflation
	gdp	forecast	inflation	terms_trade	Euro_zone	drop	portfolio	target
	expansion	expected	food	deficit	region	monetary_ policy	firms	Inflation_ expectation
	performance	colombian_ economy	variation	imports	emerging_eco nomies	fed	resources	core_ inflation
	investment	inflation	consumer_ inflation	oil	china	rates	financing	result
	dynamism	projection	tradable	oil_price	economy	increase	continue	staff
	sector	economy	regulated	observed	latin_america	market	housing_ price	indicator
	lower	predicted	increase	decrease	recover	real_ activity	debt	banco_ republica
	slowdown	economic_growth	food_price	production	risk_ premium	policy	transmission	expected
	behavior	path	adjustment	external_demand	risk	dollar	loan	level

Table A1. Topics, average weights and most relevant words in the selected LDA model for minutes and inflation reports.

Note: words in the table correspond to the ones translated from the original language (Spanish)

Appendix B

Figure B1: Topic weights within the Inflation Reports

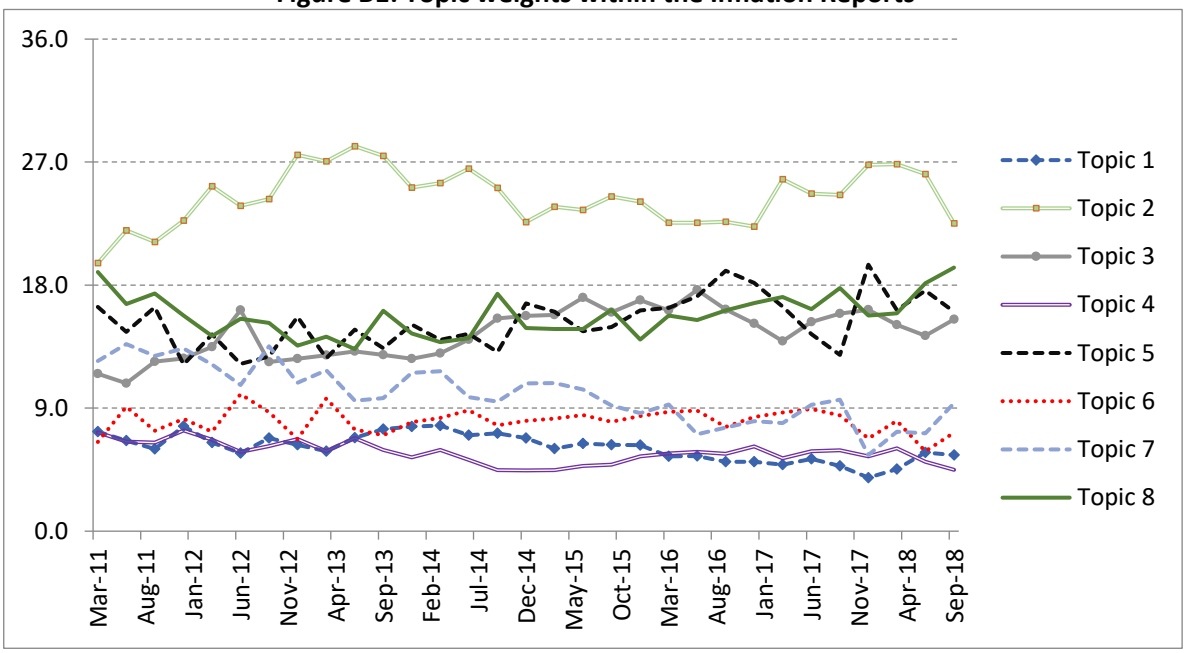
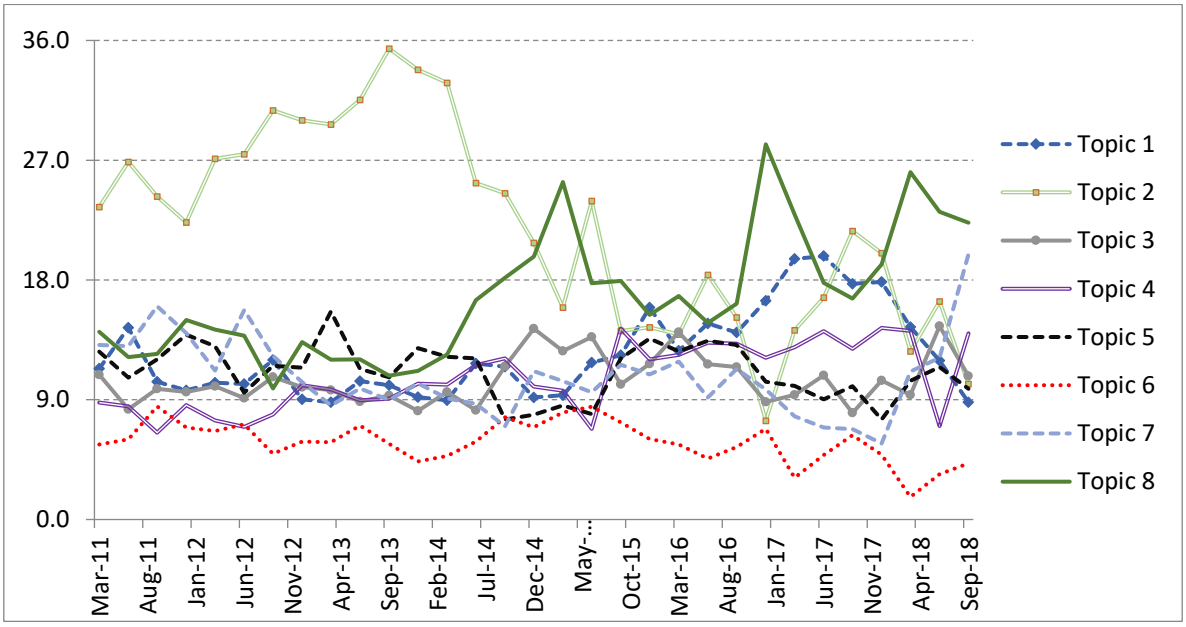


Figure B1: Topic weights within the Minutes*



* Weights correspond to the quarterly average of monthly data

Appendix C. Selection of the three most relevant narrative shocks for each maturity.

To select the key narrative shocks we conduct the following regressions:

$$\left|\Delta y_t^{rep}\right| = \gamma_{0,y} + \boldsymbol{\gamma_{1,y}} \boldsymbol{Num_t} + \gamma_{2,y} VIX_t + u_{t,y}^{rep} \quad \text{(A.1)}$$

for y = i_1, f_3, f_5 and ${}_5f_5$

$$\left|\Delta y_t^{min}\right| = \delta_{0,y} + \delta_{1,y} VIX_t + u_{t,y}^{min} \quad \text{(A.2)}$$

for y = i_1, f_3, f_5 and ${}_5f_5$

where $\left|\Delta y_t^{rep}\right|$ is the absolute value of the daily change in the market interest rate y, observed on the day of the report publication (t) and $\left|\Delta y_t^{min}\right|$ is the same variable but observed on the day of the minute publication (t).

We then use results from equations (1)-(2) and (A.1)-(A.2) to examine whether narrative shocks, $\boldsymbol{v_t}$, have an impact on the part of yields that is not explained by numeric information (u_t) and to select the three most relevant shocks:

$$\arg\min_{\boldsymbol{\beta^i}} \left\| \boldsymbol{u_y^i} - \boldsymbol{\beta_y^i v^i} \right\|_2^2 + \lambda_y^i \left[\alpha \left\| \boldsymbol{\beta_y^i} \right\|_1 + (1-\alpha) \left\| \boldsymbol{\beta_y^i} \right\|_2^2 \right] \quad \text{(A.3)}$$

for y = i_1, f_3, f_5 and ${}_5f_5$ and $i = rep, min$

Equation (A.3) corresponds to elastic net regression, the first term is associated with the objective function of the standard OLS regression and the second one is a penalty on regression coefficients, increasing in their corresponding magnitudes. We choose λ, the penalty parameter, following the same procedure described by Hansen et al. (2019b, Appendix C). Also, as suggested by those authors, we set the parameter α to 0,99 (alpha = 0 corresponds to ridge regression and alpha = 1 to the LASSO regression).

Since the elastic net regression may still select a high number of shocks, we draw a 500-simulation bootstrap procedure with replacement (as described by Hansen et al., 2019b) and record every time that each narrative shock is selected.[21] We finally choose those three shocks that are selected the highest fraction of times.

[21] A 5000-simulation bootstrap was used in the case of the inflation report.